REPORT DOCUMENTATION PAGE

OMB No 0704 0188

1. AGENCY USE ONLY (leave blank)	2. REPORT DATE 5-24-1994	3. REPORT TYPE AND DATES COVERED Interim Report

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS ONR
Statistical Methods for Minefield Detection	C N00014-94-0061

6. AUTHOR(S)

Joseph Glaz

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Connecticut Department of Statistics, U-120 196 Auditorium Rd Storrs, CT 06269-3120	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
Office of Naval Research 800 North Quincy St Arlington, Virginia 22217-5660	

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
UNLIMITED DISTRIBUTION STATEMENT A Approved for public release; Distribution Unlimited	

DTIC
SELECTED
JUL 18 1994
B

13. ABSTRACT (Maximum 200 words)

The Navy and the Marine Corps have been continually concerned about the antivehicle and antiship mines. The development of effective minefield detection procedures are of great importance as they will enhance the ability of the Navy and Marine Corps to perform their tasks. One approach that has been recently studied by the scientists of the Navy is the use of tests of randomness. In that study they express the need to develop detection methods that are based on two-dimensional processes that incorporate the dependence structure of the nearby observations. In this interim performance report four research projects related to this problem are discussed. The two-dimensional scan statistic, discussed in the last project, has the potential to be very useful in the minefield detection problem.

14. SUBJECT TERMS			15. NUMBER OF PAGES 10
Minefield Detection, Scan Statistics			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	

DISCLAIMER NOTICE

**THIS DOCUMENT IS BEST QUALITY
PRACTICABLE. THE COPY FURNISHED
TO DTIC CONTAINED A SIGNIFICANT
NUMBER OF PAGES WHICH DO NOT
REPRODUCE LEGIBLY.**

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to *stay within the lines* to meet *optical scanning requirements*.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD	- See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE	- See authorities.
NASA	- See Handbook NHB 2200.2.
NTIS	- Leave blank.

Block 12b. Distribution Code.

DOD	- Leave blank.
DOE	- Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA	- Leave blank.
NTIS	- Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

May 24, 1994

Performance Report - ONR Contract No: N00014-94-1-0061

Principal Investigator: Professor Joseph Glaz
Department of Statistics, U-120
University of Connecticut
Storrs, CT 06269-3120

DTIC QUALITY INSPECTED 8

1. Introduction.

The Navy and the Marine Corps have been continually concerned about the antivehicle and antiship mines. The development of effective minefields detection procedures are of great importance as they will enhance the ability of the Navy and the Marine Corps to perform their tasks. One approach that has been recently studied by the scientists of the Navy is the use of tests for randomness (Muise and Smith 1992). In this study they express the need to develop detection methods that are based on two-dimensional processes that incorporate the dependence structure of the nearby observations. The research supported by this contract addresses this need.

In Section 2 four research projects will be described. One research project was completed and the article based on it accepted for publication in a leading journal of statistics, Journal of the American Statistical Association, the second

94 7 15 034

94-22405

94 6 6 080

project is close to completion and will be submitted for publication by September 1. The other two project are at an early stage, but with a clear plan of solving the problems. Two graduate students have been partially supported under this contract to work on the four research projects referred above. Also, I am planning to be in touch with the Navy scientists interested in the minefield detection problem.

2. Research Projects.

Let $X_1, \ldots, X_n$ be n observations in a two dimensional rectangular region $R = [0, L_1] \times [0, L_2]$. We are interested in studying statistical procedures that are effective in detecting patterns or clusters of points. In what follows the four research projects mentioned above are described in detail.

(i) Simultaneous Confidence Intervals for Multinomial Proportions.

Assume that the rectangular field is divided as follows into two dimensional rectangular subregions: let $h_i = L_i/n_i$, $i=1,2$, where n_i are given positive integers. For $1 \le i \le n_1$ and $1 \le j \le n_2$ Define the random variables

Y_{ij} = the number of X_i's in $[(i-1)h_1, ih_1] \times [(j-1)h_2, jh_2]$.

If we assume *complete spatial randomness* (csr), i.e. that the n points are generated by a homogeneous Poisson process in the two dimensional space, then conditional on the occurrence of the n points in the region R, (in the time interval $(0,t]$) Y_{ij}, $1 \le i \le n_1$ and $1 \le j \le n_2$, have a multinomial distribution with parameters n and $p_{ij} = 1/(n_1 n_2)$. If the exact location of the points is unknown and the rectangular subregions defined above are of a considerable size, then one might be interested to examine the assumption of csr based on the observed data Y_{ij}. Assuming that the Y_{ij} are distributed according to a multinomial distribution, but possibly with not equal p_{ij}'s, we can construct a confidence region for the p_{ij}'s with a specified probability of coverage and see if the vector $(1/(n_1 n_2), \ldots , 1/(n_1 n_2))$ is in that region. Now, if csr is not warranted, then we might be willing to assume that the Y_{ij} are distributed according to a multinomial distribution, with distinct p_{ij}'s. In that case too we are interested to estimate the pij's based on the observed data.

The enclosed article "Simultanous Confidence Intervals and Sample Size Determination for Multinomial Proportions" studies this problem that has been studied by many researchers. Using a parametric bootstrap approach along with Edgeworth expansion for approximating the rectangular multinomial probabilities accurate

confidence regions are constructed for multinomial proportions. Utilizing this method we present an algorithm for determine the sample size needed to construct a confidence region with a specified volume and probability of coverage.

(ii) **Parametric Bootstrap Inference Based on Extreme Order Statistics in a Multiniomial Experiment.**

In this project we have the same set up as in (i). We assuming that the Y_{ij} are distributed according to a multinomial distribution, but possibly with not equal p_{ij}'s. Here testing the null hypotheses of car is equivalent to testing that the p_{ij}'s are equal. We propose to study the performance of the test statistic based on $\max(Y_{ij})$ and $\min(Y_{ij})$, i.e. reject H_0 if $\max(Y_{ij}) \geq a$ or $\min(Y_{ij}) \leq b$. The approximation for multinomial rectangular probabilities derived in project (i) (Sison and Glaz 1994) will be utilized in studying the performance of this test statistic. It will enable us to evaluate p-values and the power of the test. Moreover, confidence intervals for the largest and smallest multinomial proportion have been also studied. This project will be completed by September 1.

(iii) The Two dimensional Ratchet Scan Statistic

Let $X_1, \ldots, X_n$ be n observations in a two dimensional rectangular region $R = [0, L_1] \times [0, L_2]$. Let $h_i = L_i/n_i$, $i=1,2$, where n_i are given positive integers. For $1 \leq i \leq n_1$ and $1 \leq j \leq n_2$ Define the random variables

$Y_{ij} =$ the number of X_i's in $[(i-1)h_1, ih_1] \times [(j-1)h_2, jh_2]$.

If we assume *complete spatial randomness* (csr), i.e. that the n points are generated by a homogeneous Poisson process in the two dimensional space, then conditional on the occurrence of the n points in the region R, (in the time interval $(0,t]$) Y_{ij}, $1 \leq i \leq n_1$ and $1 \leq j \leq n_2$, have a multinomial distribution with parameters n and $p_{ij} = 1/(n_1 n_2)$. If we do not know the exact location of occurrence of the n observations and the rectangular regions are small we might be interested in examining the number of event that occur in several adjacent subregions. For $1 \leq i_1 \leq n_1 - m_1 + 1$ and $1 \leq j_1 \leq n_2 - m_2 + 1$ we define

$$W(i_1, j_2) = \sum_{j=i_2}^{i_2+m_2-1} \sum_{i=i_1}^{i_1+m_1-1} Y_{ij}.$$

$W(i_1, i_2)$ is the number of observations in a rectangular region
that is comprised from $m_1 m_2$ adjacent subregions. If $W(i_1, i_2)$
exceeds the value k, then k points are clustered within this
rectangular region. The two dimensional ratchet scan statistic is
defined to be

$$W_m = \max \ (W(i_1, i_2); \ 1 \le i_1 \le n_1 - m_1 + 1, \ 1 \le i_2 \le n_2 - m_2 + 1),$$

where $m = (m_1, m_2)'$. (The one dimensional ratchet scan statistic
has been recently studied by Krauth 1992). We have started
studying approximations for the distribution of W_m. The methods
that we are using include probability inequalities (Glaz 1993)
and compound Poisson approximations (Glaz, Naus, Wallenstein and
Roos 1994).

For $1 \le i_1 \le n_1 - m_1 + 1$ and $1 \le i_2 \le n_2 - m_2 + 1$ define the indicator random
variables

$$I(i_1, i_2) = 1, \quad \text{if} \quad W(i_1, i_2) \ge k$$
$$= 0, \text{ otherwise.}$$

The *multiple two dimensional ratchet scan statistic* is defined as

$$\zeta = \sum_{j=i_2}^{i_2+m_2-1} \sum_{i=i_1}^{i_1+m_1-1} I(i_1, i_2).$$

It counts the number of rectangular regions given by
$[i_1, i_1+m_1-1]x[i_2, i_2+m_2-1]$ that contain k or more observations.
Even in the one dimensional case, there are no accurate
approximations for the distribution of ζ. In this case we are
planning to employ a compound Poisson approximations (Roos 1993,
Glaz, Naus, Roos and Wallenstein 1994). We will attempt to
generalize these results in the case where the Y_{ij}'s have a
multinomial distribution but with different p_{ij}'s.

(iv) Two Dimensional Discrete Scan Statistic

In this project we have the same set up as in (iii), only
here we assume that the Y_{ij}'s are independent integer valued
random variables. The binomial and the Poisson models are
included in this study. For the special case of Y_{ij}'s being iid
Bernoulli trials the strong laws and the extreme value asymptotic
distribution of W_m has been studied in Darling and Waterman (1985
and 1986).

Under the null hypotheses of randomness we will assume that
the Y_{ij}'s are iid according to a specified discrete distribution,
for example the Poisson distribution with parameter λ_0. Under the
alternative of clustering we assume that somewhere in the
rectangular region there exists a subregion $[i_1, i_1+m_1-1]x[i_2, i_2+m_2-1]$
so that the distribution of the number of points

in that subregion is Poisson with mean $m_1 m_2 \lambda_1$, where $\lambda_1 > \lambda_0$. Then the null hypothesis of randomness is rejected at the α level if $W_m \geq k$, where $\alpha = P(W_m \geq k)$. To implement this cluster detection procedure, accurate approximations for the distribution of the two dimensional discrete scan statistic have to be derived. First, a product-type approximation similar to the one for the one dimensional case in Glaz and Naus (1991, 2.28) will be derived. Moreover, inequalities for the tail probabilities of the two dimensional discrete scan statistic will be studied. This will extend the results in Glaz and Naus (1991, Theorem 1). In the one dimensional case the bounds were asymptotically tight. It will interesting to investigate the asymptotic properties of the inequalities in the two dimensional case.

Another approach to approximate the distribution of W_m that will be studied is the Poisson and compound Poisson approximations using the Chen-Stein method (Arratia, Goldstein and Gordon 1989 and 1990, Barbour, Chen, and Loh 1991, Barbour, Holst and Janson 1992, Roos 1993, Glaz, Naus, Wallenstein and Roos 1994)). These approximations will be of great importance in investigating the distribution of the multiple discrete scan statistic that was discussed in detail in (iii) for the ratchet scan statistic.

References

Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen-Stein method. *Ann. Probab.* 17, 9-25.

Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen-Stein method. *Statistical Science* 5, 403-434.

Barbour, A. D., Chen, L. H. Y. and Loh, W. L. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* (in press).

Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford University Press, Oxford.

Chen, L. H. Y. (1975). Poisson approximations for dependent trials. *Ann. Probab.* 3, 534-545.

Cressie, N. (1984). Using the scan statistic to test uniformity. *Colloq. Math. Soc. Janos Bolayai* 45, Goodness of Fit, Debrecen, Hungary, p. 87-100.

Darling, R.W.R. and Waterman, M.S. (1985). Matching rectangles in d-dimensions: Algorithms and laws of large numbers. *Adv. Math.*, 55, 1-12.

Darling, R.W.R. and Waterman, M.S. (1986). Extreme value distribution for the largest cube in a random lattice. *Siam J. of App. Math.*, 46, 118-132.

Glaz, J. (1993). Extreme order statistics for a sequence of dependent random variables. *Stochastic Inequalities, IMS Lecture Notes - Monograph Series*, Hayward California (in press)

Glaz, J. and Naus, J. (1983). Multiple clusters on the line. *Commun. Statist. - Theor. Meth.* 12, 1961-1986.

Glaz, J. and Naus, J. (1991). Tight bounds and approximations for scan statistic probabilities for discrete data. *Ann. Appl. Probab.* 1, 306-318.

Glaz, J., Naus, J. I., Roos, M. and Wallenstein, S. R. (1993). Poisson approximations for the distribution of ordered m-spacings. Submitted for publication special issue in the *J. Appl. Probab.*

Krauth, J. (1992). Bounds for the upper-tail probabilites of the circular ratchet scan statistic. *Biometrics* 48, 1177-1185.

Muise, R. R. and Smith, C. M. (1991). Nonparametric minefield detection and localization. *Technical Memorandum NSCSS-TM-591-91*, Coastal Systems Station, Dahlgren Division Naval Surface Warfare System, Panama City, Florida.

Muise, R. (1993). A maximum likelihood algorithm for parametric minefield estimation. Coastal Systems Station, Panama City, Florida. Preprint

Naus, J.I. (1965a). Clustering of random points in two dimensions. *Biometrika*, 52, 263-267 (1965).

Sison, C. P. and Glaz, J. (1994). Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association* (in press).

Roos, M. (1993). Compound Poisson approximations for the numbers of extreme spacings. *Adv. Appl. Probab.* (in press).

Smith, C. M. (1991). Two dimensional minefield simulation. *NCSC-TM-558-91*, Naval Coastal Systems Center.